Emologus - A Compositional Model of Emotion Detection based on the Propositionnal Content of Spoken Utterances

Marc Le Tallec, Jeanne Villeneau, Jean-Yves Antoine, A. Savary, Arielle Syssau-Vaccarella

EMOLOGUS - A Compositional Model of Emotion Detection based on the Propositionnal Content of Spoken Utterances

Marc Le Tallec (a), Jeanne Villaneau (b), Jean-Yves Antoine (a), Agata Savary (a-d), and Arielle Syssau (c)
{marc.letallec,jean-yves.antoine,agata.savary}@univ-tours.fr,
jeanne.villaneau@univ-ubs.fr,arielle.syssau@univ-montp3.fr.

(a) LI, Université de Tours, France,
(b) VALORIA, Université de Bretagne Sud, France,
(c) Université de Montpellier 3, France.
(d) Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland

Abstract. The ANR EmotiRob project aims at detecting emotions in an original application context: realizing an emotional companion robot for weakened children. This paper presents a system which aims at characterizing emotions by only considering the linguistic content of utterances. It is based on the assumption of compositionality: simple lexical words have an intrinsic emotional value, while verbal and adjectival predicates act as a function on the emotional values of their arguments. The paper describes the semantic component of the system, the algorithm of compositional computation of the emotion value and the lexical emotional norm used by this algorithm. A quantitative and qualitative analysis of the differences between system outputs and expert annotations is given, which shows satisfactory results, with the right detection of emotional valency in 90% of the test utterances.

Key words: Detection of Emotions, Child SLU, Emotional Norm.

1 Introduction

A new important field of study in robotics is the domain of companion robots which execute complex tasks and offer behavior enrichment through their interaction with human beings. The French project EmotiRob, supported by the ANR (National Agency of Research), belongs to this research domain and aims at conceiving and realizing a "reactive" autonomous soft toy robot, which can emotionally interact with children weakened by disease, and give them some comfort. Previous experiments have shown the contribution of companion robots in this type of situation.

In the EmotiRob project, the robot simulates emotional states by facial expressions. To enable it to simulate a pertinent emotion, our system aims at detecting emotions conveyed in words used by children by combining prosodic and linguistic clues. Only the latter ones are addressed in our present study, i.e., we rely merely on the propositional content of a child's uttenrance.

2 Detection of emotions

There is currently no consensus about what an emotion is and how an emotion has to be characterized. An emotion is a complex cognitive state, which is strongly dependent on various contexts: short-term context includes the type and the circumstances of the interaction, while long-term context is related to cultural and personal life. We resume the two approaches which are most used to characterize emotions. In the first one, emotions are classified into emotional modalities. The set of modalities may vary but most of authors agree to a classification into seven emotional modalities: anger, disgust, enjoyment, fear, surprise, sadness and neutrallity [4, 3]. The second approach uses an ordinal classification in a muldimensional space. For example, some psycholinguists use excitement level and emotional valency (negative/positive). All these works show both of the following conclusions:

1. In a real dialogue, most of the speech turns do not convey perceptible emotion, as 80% of them are classified as neutral from an emotional point of view.
2. The perception of the emotions is very variable. The measures of the inter-annotator agreement give poor results. A referential annotation may be achieved with a majority vote only.

Most of the time, emotion detection performs classifications, by using acoustic or prosodic clues. The use of linguistic clues such as indications of repairs or presence of emotional words is not frequent, although this use improves the performances of the systems [8]. Besides, these performances are still perfectible.

Because linguistic emotion detection seems a hard task and has hardly been investigated, we have chosen to represent emotion as a simple pair (valency, intensity), where valency can be negative, positive or neutral and where intensity is measured by an integer from 0 to 2.

Before realizing an automatic system of emotion detection based on those principles, it was necessary to know if the task had a chance of success, Therefore we first tested how good the agreement of annotators can be on emotions conveyed by the lexical content of sentences produced by children. Choosing a representative test corpus is not easy, since an interactive emotional robot for children does not exist currently. We collected a corpus in a primary school, where about 7-year-old children were asked to invent tales. The corpus (so-called Brassens corpus) is composed of about 170 sentences which make up twenty short stories. Two annotations were performed by nine (5 adults and 4 children) annotators: in the first one, sentences were given in a random order, so that an out-of-context annotation can be obtained, while in the other one the sentences were given in the order of the stories. Agreement between annotators is calculated by using Kappa coefficients, which show low correlation between the child annotators (0.49 for the out-of-context and 0.38 for the contextual annotation), however the various ages (from 4 to 9) of the four annotators can partially explain these poor results. On the other hand, there is a good agreement between adult

annotators (0.86 for the out-of-context and 0.84 for the contextual annotation), an encouraging result within the framework of our purpose.

3 Natural Language Understanding

Our objective is to detect emotions which are conveyed in the propositional content of a linguistic message. One can presume - and this assumption has been confirmed by our surveys (see Section 5) - that many words convey a positive or negative emotion by themselves. Therefore, a simple first approach to obtain an emotional measure of a message is to add up the emotional measure of each of its words. An advantage of this solution is that no understanding of the message is required. An evaluation of the emotional potential of each domain word has to be achieved only. Our baseline is based on this principle.

Nevertheless, it is obvious that this solution does not work in all cases. For example, *"la mort de la méchante sorcière" (the dead of the mean witch)* does not convey a negative emotion although each of its words does. In this example, it can be assumed that the emotional potential of the concept *"death"* is dependent on its related object. Based on these principles, our work aims at realizing a compositional calculus of the emotional content of the utterances. To achieve that, a semantic treatment of the utterances is required, which achieves their "understanding" i.e. specifying the semantic linkages between concepts.

Spoken Language Understanding (SLU) is a difficult task, especially because of speech recognition errors and spoken disfluencies. A very robust parsing is required and current operational systems are related to restricted tasks with a restricted vocabulary [9, 5, 14]. Thus, the objective of trying to understanding the utterances of children within the framework of the EmotiRob project may seem unattainable. To make the task feasible, the domain has been restricted to the concepts of the world of very young children, 4 or 5 years old. Moreover, complete understanding is not always necessary: the baseline may replace predicative calculus in the case of partial understanding.

Our SLU system is based on logical formalisms and performs an incremental deep parsing [12]. It provides a logical formula to represent the meaning of the word list that Automatic Speech Recognition provides to the SLU as input. The understanding module performs a translation from natural language to a target logical formalism. The vocabulary known by the system as the source langage contains about 8000 lemmas selected from the lexical Manulex[1] and Novlex[2] bases. We have restricted the concepts of the target langage by using Bassano's studies related to the development of child language [2]. SLU carries out a projection from the source langage into Bassano's vocabulary.

The parsing is split into three main steps: the first step is chunking [1] which segments the sentence into minimal syntactic and semantic groups. The second step builds semantic relations between the resulting chunks and the third is a contextual interpretation. The second and third step use a semantic knowledge

[1] http://leadserv.u-bourgogne.fr/bases/manulex/manulexbase/indexFR.htm
[2] http://www2.mshs.univ-poitiers.fr/novlex/

of the application domain. Thus, the main work that had to be done to adapt the system to our objective was to build an ontology from the set of application concepts, a difficult task due to the width of the application domain. More precisely, Bassano vocabulary includes many verbs, some of them with polysemic meaning. To specify the possible uses of these verbs, a part of the ontology [6] is based on a linguistic corpora study related to fairy tales.

Figure 1 shows the parsing of the utterance: *"Il était une fois un petit cochon qui n'avait pas d'amis" (Once upon a time there was a little pig who had no friends).* Chunking provides six chunks which are gradually linked in the following parsing steps. The logical formula provided by the system is:

(narrative (neg (to_have [(subject: (pig [(size: little)])), (object: (friends))])))

The calculus related to emotion detection of this utterance is given in the following sections.

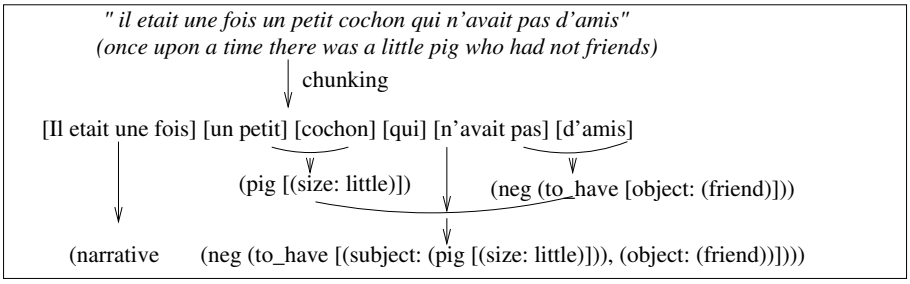

Fig. 1. An example of parsing.

4 Basic principles of the EMOLOGUS system

In the EMOLOGUS system, the detection of emotions relies on a major principle: the emotion conveyed by an utterance is compositional. It depends on the emotion of every individual word as well as the semantic relations characterized by the SLU system. More precisely, simple lexical words have an intrinsic emotional value, while verbal and adjectival predicates act as a function on the emotional values of their arguments. As an illustration, consider the sentence of Fig. 1 and its related logical formula. The computation of the emotion begins with the consideration of the emotional value of the words *pig* and *friends* ($E = 0$ for *pig* and $E = 1$ for *friends*), which are simple arguments of the formula. Then, adjective such as *little* and verbs such as *to have* acts as predicate on these initial values. For instance, *little pig* is assigned $E = +1$ as emotional value, since *little* is defined as the predicate *E:* $x \mapsto x + 1$. The successive applications of the predicates provide the global emotional value of the sentence: $E = -1$ (Fig. 2). As an illustration, here are different definitions of predicates found in our lexicon:

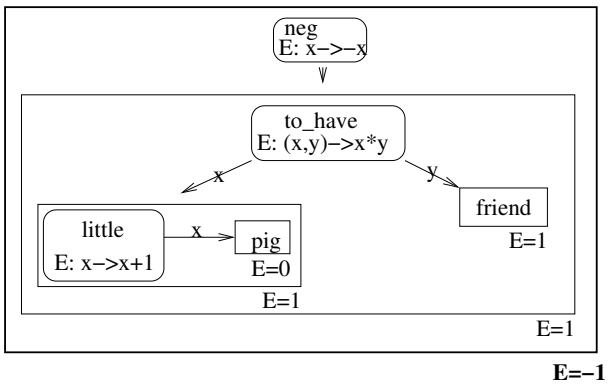

Fig. 2. Compositional calculus for the example sentence (cf. Fig. 1).

$E : x \mapsto x + 1$ (aimable / kind) $E : x \mapsto x - 1$ (énervé / irritated)
$E : (x, y) \mapsto -y$ (casser / to break) $E : (x, y) \mapsto 1$ (chatouiller / to tickle)
$E : (x, y) \mapsto min(x, y)$ (accompagner / to go with)

5 Emotional norm and definition of predicates

The first requirement is to know emotional valency which is associated to the lexicon words by children. This has been the aim of emotional lexical standards which have been used for a long time in experimental psychology. These standards compile subjective evaluations of a population of judges about one or several emotional characteristics of words.

Some standards are related to emotional characteristics such as the duration of the emotion caused by a word [7, 13]. However in all emotional lexical standards, two characteristics are always estimated: valency and intensity.

Both characteristics are mainly estimated by an adult population, using nominal scales of judgment (positive, neutral, negative) or ordinal (i.e., -5 very negative through 5 very positive). To our knowledge, only two standards have compiled the evaluations made by young children: Vasa et all [11] for the English language, and Syssau and Monnier [10] for the French language (5, 7 and 9-year-old children). In these standards, the answer scales used are the same as those used with adults with slight modifications. The number of modalities is reduced (3 for the study of Syssau and Monnier) and every answer modality is related to a drawing which represents a smiling, sad or neutral face, respectively. The examination of the results shows that from 5 years of age, the children are able to judge emotional valency of the words with a substantial agreement.

In the EmotiRob project, we complete the standard of Syssau and Monnier by the evaluation of emotional valency of 80 new words extracted from the Bassano lexicon with children between 5 and 7 years old. In the original standard, the words were classified by age of acquisition and most of them were common nouns. For the 5-year-old children, the added words, most of them adjectival or verbal,

have been divided into 2 lists of 40 words estimated in two different sessions. For those of 7 year olds, the list was to be estimated in a single session. At every age, two random orders of presentation of the words was defined, every order being presented to half of the participants. These experiments were carried out in 4 French schools.

To complete the characterization of our lexicon, an emotional predicate has been assigned to every verb or adjective of our application lexicon through an agreement procedure among five adult experts. More precisely, every expert proposed one or at most two definitions for every predicate. Then, agreement was sought among these proposals. It is interesting to note that it has finally been possible to reach a complete agreement.

6 Experiments and Results

We conducted several experiments in order to assess the behaviour of our system. These experiments have been carried out on the Brassens corpus (cf. section 2), by using annotations of 5 experts as test references. This evaluation assessed the detection of emotion without considering the discursive context. This is why the test sentences have been provided in a random order to the annotators, who had to describe the emotion valency conveyed by a single scalar value including valency and intensity between -2 (very negative) and 2 (very positive). The reference was obtained through a majority ballot among the expert annotations. Finally, we compared the semantic Emologus system with the basic baseline presented in section 3 on 173 emotionally annotated sentences. The results are shown in the following table:

	Baseline	Emologus
Precision	68.8%	90%

With a precision of 90%, Emologus presents a good accuracy, by opposition with the baseline. This result suggests that the detection of emotions should greatly benefit from the consideration of the linguistic content, in addition to a standard prosodic analysis.

Emolog.\ Ref.=

	-2	-1	0	1	2
-2	4	2	0	0	0
-1	2	18	0	0	0
0	1	5	116	2	0
1	0	0	3	16	1
2	0	0	0	1	2

Baseline\ Ref.=

	-2	-1	0	1	2
-2	4	0	0	1	0
-1	3	12	7	0	0
0	0	6	90	4	0
1	0	4	18	11	1
2	0	0	7	3	2

Table 1. Matrix of confusion for Emologus and the baseline, respectively.

Table 1 presents the error confusion matrices of the two systems, which enable an in-depth analysis of the error distribution. Within the EmotiRob context, the

most serious error of the system is giving an opposite valency, because it can infer a bad reaction of the robot. This type of error, called "valency inversion", is never observed with EMOLOGUS. Most of its errors correspond to "emotion deletions" errors, when the system does not detect an emotion which is present in the test reference. The opposite error, "emotion insertion" only concerns 18% of the errors made by EMOLOGUS. The last kind of error is less serious: it corresponds to "intensity errors" when the system detects the correct valency but assigns an erroneous intensity (for instance: "positive" vs. "very positive"). Fortunately, 35% of the errors of EMOLOGUS are intensity errors. If one ignores such moderate errors, the precision of the system rises up to 94%. To the contrary, the baseline leads to more serious errors, among which are valency inversions.

Remarks related to the analysis of some errors

Some verbs naturally have a positive or negative emotion when we do not know who is doing the action, for example in *"La femme était enfermée dans une prison" (the lady was locked in a jail). transl. "To be locked in a jail"* is very negative when the subject is positive, and very positive when the subject is negative. When it is not possible to know who is locked up, a low negative emotion is felt by annotators. It is the same for instance for *can't do something*, *don't believe someone* and the opposite for *find something*. A solution would be to define a default emotional behaviour when the valency of an argument is unknown or neutral.

Some errors result from an erroneous modelisation of adjective or adverb of degree, as *little*. On the whole, this adjective involves a general tendency to shift the valency of its argument to more positive values, as in *a little wolf*. This is why it has been modelled by the predicate: *E:* $x \mapsto x + 1$. Unfortunately, the influence of the adjective should differ in some specific situations.

However, the human annotators do not always respect this behavior. Consider the sentence *"il a vu une petite maison" (transl. "he has seen a little house")*. Since *maison/house* does not support any emotion (E = 0), the nominal group *petite maison/little house* should be considered as slightly positive (E = 0 + 1 = 1). However, the majority of our experts consider it as neutral. This means that in some circumstances that still have to be investigated, the adjective does not affect the emotional valency of its arguments. As a result, *little* should present different emotional behaviors. This is a good example of what we can call an emotional ambiguity. Fortunately, the latter seems to be moderate.

We also have problems with elements which directly depend on the context of the story. For instance: *Les parents s'enfuirent (transl. The parents ran away)*. Here, we can't be sure if it is positive or negative. The emotional interpretationof such a sentence depends highly on the discourse context: did the parents succeeded in avoiding a danger, or did they have doing something wrong. Naturally, annotators have annotated it with a neutral emotion, while the semantic model chose a positive emotion due to the positive valency of the noun *parents*. For more precision we need more information about why they run away.

7 Conclusion and perspectives

We have tested our semantic model on a corpus, and results are encouraging. These experiments show it is possible to detect emotion on the basis of linguistic clues with a high precision (90%). A very positive fact is that we never find opposite valency. We have modified the emotional function of some predicates, because of the particularities we have found in these experiments, in particular in sentences involving emotionally neutral subjects. We have to verify these results on a larger corpus, before working on sentences in context, with management of anaphora. Also, we have to think about how to combine emotions in several sentences and about dynamics of emotion on a complete text.

References

1. Abney, S.: Parsing by Chunks. In: Principle Based Parsing. R. Berwick, S. Abney and C. Tenny, Eds., Kluwer Academic Publishers. (1991)
2. Bassano, D. et al.: Le DLPF, un nouvel outil pour l'évaluation du développement du langage de production en français. In: Enfance, 2(5):171–208. (2005)
3. Cowie R., Cornelius R.: Describing the emotional states that are expressed in speech. In: Speech Communication. 40. 5-32. (2003)
4. Ekman, P.: Patterns of emotions: New Analysis of Anxiety and Emotion. Plenum Press. (1999)
5. Glass, J.: Challenges for Spoken Dialogue Systems. In: Proceedings IEEE ASRU. Workshop, Keystone, Colorado, USA. (1999).
6. El. Maarouf, I.: Natural ontologies at work: Investigating fairy tales. In: Corpus Linguistics 2009, Liverpool (G.B.). (2009)
7. Niedenthal P. M., Auxiette C., Nugier A. et al.: A prototype analysis of the French category émotion. In: Cognition and Emotion, 18. 289-312. (2004)
8. Schuler B. et al.: The Relevance of Feature Type for the Automatic Classification of Emotional User States. In: Interspeech2007, Anvers, Belgique. 2253-2256. (2007)
9. Seneff, S. and Polifroni, J. A new restaurant Guide Conversational System: issues in Rapid Prototyping for Specialized Domains. In: International Conference on Spoken Language Processing (ICSLP'96), pages 665–668, Philadelphia. (1996)
10. Syssau A., Monnier C.: Children's emotional norms for six hundred French words. In: Behavior, Research, and Methods, 41, 213-219. (2009)
11. Vasa R. A., Carlino A. R., London K., Min C.: Valence ratings of emotional and non-emotional words in children. In: Personality and Individual Differences, 41, 1169-1180. (2006)
12. Villaneau, J. and Antoine, J-Y.: Deeper Spoken Language Understanding for Man-Machine Dialogue on Broader Application Domains: A Logical Alternative to Concept Spotting. In: Proceedings of SRSL 2009, the 2nd Workshop on Semantic Representation of Spoken Language., Athens, Greece. 50–57.
13. Zammuner V.L.: Concepts of emotion: Emotioness and dimensional rating of Italian emotion words. In: Cognition and emotion, 12, (2), 243-272. (1998)
14. Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazen, T., and Hetherington, L.: JUPITER: Telephone-Based Conversational Interface for Weather Information. In: IEEE Transactions on Speeh and Audio Processing, XX(Y):100–112. (2000)